Greg Hundermark

Mrs. Dianne Krause

GRIT 687

June 25, 2012

Automated Essay Scoring

A long-held dream for high school English teachers and college professors alike is making headlines, and the reason why may come to many as a surprise. The dream in question is Automated Essay Scoring (AES), or “a measurement technology in which computers evaluate written work” (Shermis & Burstein). While the first successful automated essay-scoring engine first appeared in 1973, limitations of the technology at the time prevented AES from seeing any practical usage (Shermis et al. 3). Fast-forward several decades later, and the Internet has revived this technology. Shermis et al. adds, “The internet provide[s] a universal platform for writers to submit text for review” (Shermis et al. 3). The accessibility of the Internet, in combination with commercial software companies offering AES to schools at affordable rates, has led to dialogue about the pros and cons of this emerging technology.

In a recent *The New York Times* article, author Randall Stross published the highlights of a competition hosted by Kaggle, a company whose goal was “to see how well algorithms submitted by professional data scientists and amateur statistics wizards could predict the scores assigned by human graders” (Stross). Similarly, Stross noted that during the same time of Kaggle’s competition, The Hewlett Foundation sponsored a study of the already existent commercial AES services. In both the competition hosted by Kaggle and the research sponsored by The Hewlett Foundation, the algorithms scored nearly identical to those of human graders (Stross).

While the thought of AES replacing humans makes many people uncomfortable, Stross is quick to counter. He writes, “In states’ standardized tests, each essay is typically scored by two human graders; machine scoring replaces only one of the two. And humans are not necessarily ideal graders: they provide an average of only three minutes of attention per essay…” (Stross). While Stross’s article reads as if it is pro-AES, he does acknowledge its shortcomings. He writes, “We are talking here about providing a very rough kind of measurement, the assignment of a single summary score on, say, a seventh grader’s essay, not commentary on the use of metaphor in a college senior’s creative writing seminar” (Stross). While Stross does not elaborate on the failures of AES to assess a college senior’s use of metaphors, other researchers have.

Les Perelman, a director of writing at the Massachusetts Institute of Technology, has had success “gaming” *E-Rater*, the only company willing to grant Perelman access to their AES software (Winerip). One problem, according to Perelman, is that E-Rater does not currently have the ability to assess substance (Winerip). For example, a student can write that the World Trade Center attacks happened on any date in history and E-Rater does not flag the inaccuracy. Another limitation of AES is that E-Rater grades longer essays more favorably than shorter essays (Winerip). Winerip writes, “A 716-word essay [Perelman] wrote that was padded with more than a dozen nonsensical sentences received a top score of 6; a well-argued, well-written essay of 567 words was scored a 5”. Similarly, E-Rater does not like anything else that is short; this includes words, sentences, and paragraphs. Instead, E-Rater prefers transitional words, as well as “big” words, as it interprets them as a sign of lexical complexity (Winerip). Given the benefits and disadvantages of AES, the conversation shifts to if and how AES can help learners in its current state.

Primarily, AES can serve as a classroom writing aid. After an instructor assigns an essay, students can write and submit their essay for scoring as many times as the instructor allows. Students can utilize the instant feedback from the algorithm to make edits to their writing and take advantage of a scenario that otherwise cannot be replicated by a classroom instructor. From my experience using *MyAccess!*, a commercial AES product, the algorithm identifies problem-areas that I would not have otherwise recognized, which results in an ideal scenario for both the learner and the instructor. Ultimately, the instructor grades a student’s final draft that is far superior to one that in a traditional setting has only been revised a handful of times, if at all.

Admittedly, on rare occasions I have used AES as a replacement for human assessment, as opposed to as the previously detailed scenario of AES only being used as a classroom writing aid. In my honors classes, I feel inclined to assign a fair number of essays, but with over sixty students, the ability to provide timely and authentic feedback for each of these essays is taxing and sometimes superficial. *MyAccess!* allows me to assign more essays than I typically would, with the opportunity for learners to request teacher intervention only if a pre-determined set of criteria has been met (i.e. Learners must prove they followed a standard format, spellchecked and proofread, and showed evidence of multiple revisions). While AES does have its shortcomings, I see this as a fair compromise for students who feel that AES is underscoring their writing.

I predict that in the future, AES will become a transparent feature of word-processing applications, just as spelling and grammar suggestions are today. By infusing this emergent technology into the age-old practice of writing, the instructor’s current role of assessment will shift. Stross adds, “Instead of the teacher filling the essay with the markings that flag problems, the software could do so. The software could also effortlessly supply full explanations and practice exercises that address the problems — and grade those, too”. Ideally then, instructors can focus more on the actual substance of the essay – like if a student erroneously writes that dinosaurs are fictional, which unfortunately I have been exposed to.

Works Cited

Shermis, Mark D., and Jill Burstein. *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Lawrence Erlbaum Associates: New Jersey, 2003. Print.

Shermis, Mark D., Jill Burstein, Derrick Higgins, and Klaus Zechner. *Automated Essay Scoring: Writing Assessment and Instruction*. Publication. N.p.: n.p., n.d. Print.

Stross, Randall. "DIGITAL DOMAIN; The Algorithm Didn't Like My Essay." *The New York Times*. The New York Times, 10 June 2012. Web. 25 June 2012. <http://www.nytimes.com/2012/06/10/business/essay-grading-software-as-teachers-aide-digital-domain.html?\_r=1>.

Winerip, Michael. "ON EDUCATION; Facing a Robo-Grader? No Worries. Just Keep Obfuscating Mellifluously." *The New York Times*. The New York Times, 23 Apr. 2012. Web. 25 June 2012. <http://www.nytimes.com/2012/04/23/education/robo-readers-used-to-grade-test-essays.html?pagewanted=all>.